DOCUMENT RESUME

ED 326 539 TM 015 183

AUTHOR Stivers, Janet
TITLE Relating the NTE to Research Literature on Teaching Effectiveness.
PUB DATE Apr 90
NOTE 36p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (Boston, MA, April 17-19, 1990).
PUB TYPE Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Administrators; *College Faculty; Delphi Technique; *Educational Research; Higher Education; *Instructional Effectiveness; Licensing Examinations (Professions); Literature Reviews; *Research Utilization; Teacher Certification; Teacher Educators; *Teacher Effectiveness; Test Items; Test Reliability; Test Validity
IDENTIFIERS *National Teacher Examinations; Subject Content Knowledge; *Test of Professional Knowledge

ABSTRACT

Findings of teaching effectiveness research were related to the National Teacher Examinations (NTE) Test of Professional Knowledge (TPK). Delphi methodology was used to classify TPK items according to findings in a selected review of research. Six professors in teacher education completed a questionnaire to rate seven literature reviews. The three most highly related were presented to the Delphi panelists--three public school administrators and six college teachers from various areas of teacher education. In Phase 2, two panels of six experienced classroom teachers each considered half of the 104 TPK items in light of the selected review. Of the 104 TPK items, 20 were classified as supported by research; only 9 of the 20 were judged to require knowledge of that research. It is suggested that the TPK may be missing opportunities to measure important aspects of teachers' professional knowledge. Alternate strategies for measuring knowledge of teaching effectiveness research are considered for the planned successor to the NTE. Four tables contain study data. A 51-item list of references is included. (SLD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED326539

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

JANET STIVERS

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Relating the NTE to Research Literature on Teaching Effectiveness

Janet Stivers, Marist College

Abstract

The purpose of this study was to relate findings from teaching effectiveness research to the NTE Test of Professional Knowledge (TPK). Delphi methodology was used to classify TPK items according to findings in a selected review of research. Of the 104 TPK items, 20 were classified supported by research; only nine of the 20 were judged to require knowledge of that research. One conclusion is that the TPK may be viewed as missing opportunities to measure important aspects of teachers' professional knowledge. Alternate strategies for measuring knowledge of teaching effectiveness research are considered for the planned successor to the NTE.

Paper presented at the annual meeting of the
National Council on Measurement in Education

Boston
April, 1990

TM015183

Of the many education reforms recommended in recent years, few have been implemented as quickly as those calling for teacher testing. Before 1980, only three states required teacher applicants to pass initial teacher certification tests; currently, 44 states have laws or regulations requiring such testing. The test that is used most often is the NTE (previously known as the National Teacher Examinations), currently used in 22 states (Rudner, 1987). Because the testing of teachers is a widespread and apparently stable phenomenon, it is important to insure that, among other things, the tests that are used are well constructed.

Determining the appropriate content for a test of teachers' professional knowledge, particularly one to be used nation-wide, presents certain difficulties because of a lack of agreement about the appropriate content of teacher education curricula, which have been described as fragmented and unstable. Although there is disagreement about what should be taught in teacher preparation programs, few would fail to accord a place in the curriculum to conclusions drawn from research on teaching. Correlational and experimental research conducted over the past twenty years has produced a body of knowledge with important implications for teaching practice (Berliner, 1984; Gage, 1978, 1985; Good, 1983; Hosford, 1984; Hunter, 1984; B. O. Smith, 1985; D. C. Smith, 1982, 1983). Several experimental studies have demonstrated that teachers can learn to use recommendations drawn from research, and that those who do so have students who make greater achievement gains than do students of other teachers (e.g., Anderson, Evertson, and Brophy, 1979; Borg and Ascione, 1982; Crawford, Gage, Corno, Stayrook, and Mitman, 1978; Emmer, Sanford, Evertson, Clemens, and Martin, 1981; Good and Grouws, 1979, 1981; Emmer, Sanford, Clemens, and

Martin, 1982; Stallings, Needels, and Stayrook, 1979; cited in Gage, 1985). Another experimental study has demonstrated that, like practicing teachers, preservice teachers can learn research-based techniques and apply them in the classroom (Hindman and Polsgrove, 1988). Several researchers and teacher educators have emphasized the importance of incorporating knowledge derived from what is commonly called "teacher effectiveness research" into the content of teacher education programs (Berliner, 1984; Clark, 1984; Doyle, 1982; Egbert, 1984; Gage, 1978, 1985; Good, 1983; Hersh, 1982; Kluender, 1984; National Commission for Excellence in Teacher Education, 1985; B. O. Smith, 1985; Stallings, 1984.) Presumably, this knowledge also should be represented on initial teacher certification tests.

It is particularly important to verify that findings from research on teacher effectiveness, such as those identified by Brophy and Good (1986) and Berliner (1984), are represented on the Test of Professional Knowledge because Educational Testing Service (ETS), administrator of the NTE, has assumed a role in shaping teacher education curricula. ETS has approved the use of the NTE for the evaluation of teacher education programs and offers item summary workshops to help college personnel identify curricular areas that might be modified to improve their students' test performance. Further, because the test is used for selection, the emphases in the test can be viewed by education faculty and students as crucial, necessary, and appropriate targets for instruction. Thus, the NTE can be viewed as shaping the definition of "professional knowledge" for teachers.

The purpose of this study was to estimate the extent to which the findings from research on teacher effectiveness are represented on the NTE Test of Professional Knowledge. The specific purposes were:

1. to select, from several reviews of research on teacher effectiveness, those that are most reflective of the findings which should

be included on an initial teacher certification test,

2. to identify research findings that are judged to be represented among the items on a released form of the TPK, and

3. to identify items that are judged to be supported by research findings included in the selected review, and for items judged supported by research, to identify those items for which knowledge of research findings is important in selecting the keyed option; also, to identify items that are judged to be contradictory to research findings, that is, items for which the research literature might suggest a different keyed answer.

The study was conducted in two phases. In Phase I, a preliminary screening followed by a two-stage Delphi investigation was used to identify relevant reviews of research; in Phase II, a two-stage Delphi process was used to examine TPK items in light of the research findings included in the review(s) selected in Phase I.

Phase I

Preliminary Screening

Research on teacher effectiveness has been reviewed and summarized by several authors (e.g., Berliner, 1984; Brophy and Good, 1986; Doyle, 1986; Rosenshine and Stevens, 1986; United States Department of Education, 1986, 1987). To identify the reviews that would be most appropriate for use in this study, a preliminary screening was conducted. First, a list of reviews was compiled which included all those review articles (rather than book-length works) which had been published within the previous five years, which reviewed several studies rather than describing a single research project, and which discussed research findings applicable to a variety of educational settings.

Twenty university professors in teacher education, instructional psychology, and educational psychology were then contacted by letter and asked if they would be willing to rate such reviews on several dimensions. They were also asked to examine the list of reviews that might be included on a rating form, and to suggest other appropriate reviews. Fifteen professors responded, with ten agreeing to participate and three of the ten suggesting changes that resulted in two substitutions and two additions to the list of reviews. The reviews rated were:

Berliner, D. C. (1984). The half-full glass: A review of research on teaching. In P. L. Hosford (Ed.), Using what we know about teaching (pp. 51-77). Alexandria, VA: ASCD.

Brophy, J., & Good, T.L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 328-375). New York: Macmillan.

Doyle, W. (1986). Classroom organization and management. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 392-431). New York: Macmillan.

Rosenshine, B. & Stevens, R. (1986). Teaching functions. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 376-391). New York: Macmillan.

U.S. Department of Education. (1987). What works: Research about teaching and learning. (2nd ed.). Washington, D.C.: Author.

Walberg, H.J. (1986). Syntheses of research on teaching. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 214-229). New York: Macmillan.

Wyne, M. D. & Stuck, G. B. (1982). Time and learning: Implications for the classroom teacher. Elementary School Journal, 83, 68-75.

Results. Six professors completed a single questionnaire, using a scale of 1--poor to 5--superior, to rate the reviews on each of five dimensions: scholarship, comprehensiveness, understandability, freedom from reviewer bias, and emphasis on general rather than grade- or subject-specific findings. Their ratings and the means are presented in Table 1.

Variability among the raters is evident. For example, Professor C

TABLE 1

RESULTS OF PRELIMINARY SCREENING: MEANS & RATINGS

	Scholarship	Comprehen-siveness	Understand-ability	Freedom From Bias	Emphasis on General Findings	Mean/ Sum
Berliner, D. C. (1984)	3.25 3 - - 4 4 2	2.75 2 - - 4 3 2	4.00 4 - - 4 4 4	3.25 3 - - 4 4 2	3.50 4 - - 4 4 2	3.35 67
Brophy & Good (1986)	4.50 5 5 5 5 4 3	4.50 4 5 5 5 5 3	4.00 3 5 5 4 3 4	4.17 5 5 5 5 3 2	4.00 4 5 5 5 3 2	4.23 127
Doyle, W. (1986)	4.17 4 4 5 5 4 3	4.00 3 3 5 5 5 3	4.17 3 4 5 5 4 4	4.17 4 4 5 5 4 3	3.83 3 4 5 5 4 2	4.07 122
Rosenshine & Stevens (1986)	4.00 3 5 5 4 4 3	4.00 4 5 5 3 4 3	4.67 5 5 5 4 5 4	3.83 4 5 5 3 4 2	3.75 4 5 5 2.5 4 2	4.05 121.5
U.S. Department of Education (1987)	2.17 2 2 2 3 3 1	3.17 5 2 3 4 3 2	4.00 5 3 3 4 5 4	2.17 2 2 2 3 3 1	3.67 5 3 3 4 4 3	3.03 91
Walberg, H. J. (1986)	4.33 3 5 5 5 5 3	3.67 4 3 4 4 4 3	3.00 2 3 3 4 4 2	3.83 3 5 5 3 4 3	3.83 4 5 5 4 3 2	3.73 112
Wyne & Stuck (1982)	3.00 3 - - 3 4 2	2.50 2 - - 3 3 2	3.50 3 - - 4 3 4	3.50 4 - - 4 4 2	2.75 3 - - 4 3 1	3.05 61

NOTE: Raters used the following scale: 5- superior; 4 - above average; 3 - average; 2 - below average; 1 - poor

The ratings are presented in order, i.e., the ratings assigned by panelist A are always in the first position and panelist C's ratings are in the third position, etc.

A " - " indicates that a panelist chose not to rate a particular dimension or review.

7

8

rated five reviews, and gave three of them "perfect scores," assigning 5s on each dimension. By contrast, in rating seven reviews, Professor F assigned no 5s at all, and assigned 4s on only one dimension (understandability). Within the 35 cells of the questionnaire (seven reviews, each rated on five dimensions), there were only three instances in which a panelist assigned a rating lower than that assigned by Professor F.

Despite the variability, there are consistencies. Each professor assigned his or her lowest ratings to the United States Department of Education publication, What Works (1987), on the dimensions of scholarship and freedom from reviewer bias. No one assigned ratings of 5 to the reviews by Berliner (1984) or Wyne and Stuck (1982). If one were to generate for each panelist a rank order list of the reviews, (produced by summing the panelist's ratings across all the dimensions of the review), the article by Wyne and Stuck would occupy the lowest position on each of the lists on which it appears.

Only three reviews had mean ratings above 4 on the scale of 1 to 5; the same three reviews also represented all of those rated highest and second-highest by each professor. Therefore, the reviews selected for use in the next stage of Phase I were Brophy and Good's (1986) "Teacher Behavior and Student Achievement", Doyle's (1986) "Classroom Organization and Management", and Rosenshine and Stevens' (1986) "Teaching Functions".

Delphi Investigation

Panelists. Individuals participating in the Delphi portion (Linstone and Turoff, 1975) of Phase I were asked to examine the reviews of research by Brophy and Good, Doyle, and Rosenshine and Stevens, and to rate each in terms of how fully it reflected the research which should be

included on a paper-and-pencil test for the intial certification of teachers. This task required the judgement of experts; individuals invited to participate in this portion of the study qualified as experts on the basis of their academic credentials, experience with beginning teachers, and knowledge of research on teaching. Each of the panelists had an earned doctorate and five or more years experience supervising beginning teachers, had read two or more reviews of teacher effectiveness research, and was familiar with the work of five or more researchers often cited in the reviews. In addition, to insure that the panel members represented a variety of disciplines and were likely to have expertise in different aspects of the teacher effectiveness research literature, the assembled panel included specialists in: elementary reading, elementary social studies, secondary English, secondary math, special education at the elementary and secondary levels, and supervision and staff development at the elementary, secondary, and district levels. Three panelists were public school administrators: a principal, an assistant superintendent for instruction, and an acting superintendent. The remainder were college professors teaching content and/or methods courses. All had teaching experience at the elementary and/or secondary level.

Procedure. Panelists were mailed copies of the reviews by Brophy and Good, Doyle, and Rosenshine and Stevens, and asked to examine and rate each on a scale of 1 to 7, from minimally to highly reflective of the research on teacher effectiveness which should be included on a paper-and-pencil test used in the initial certification of teachers. Panelists also were asked to supplement their ratings with comments.

Ratings and comments were then returned to the Delphi coordinator, who tabulated them and returned to each panelist a summary sheet listing, for each review of research, the mean, median, and the first and third

quartiles for the ratings, and all comments made by panelists about the review. Panelists were then asked to rate the reviews again in light of this feedback. If a panelist's second rating fell below the 25th percentile or above the 75th percentile, he or she was asked to state the reason for assigning a rating that was markedly different from the judgment of the group. The ratings and comments produced in this second round were then forwarded to the Delphi coordinator for tabulation.

Because a consensus emerged from the second-round ratings, a third round was unnecessary.

Results. The first-and second-round ratings of each panelist are presented in Table 2. As is most often the case in Delphi investigations, panelists were generally responsive to the feedback and tended to change their ratings in the direction of the perceived consensus.

The highest final ratings were assigned to the review by Brophy and Good, which was rated above the midpoint by 8 of the 9 panelists. These ratings contrast sharply with those for the review by Doyle, which was rated below the midpoint by 8 of the 9 panelists, and contrast moderately with the ratings for the Rosenshine and Stevens review, which received 5 ratings below the midpoint, 3 ratings at the midpoint, and one rating above the midpoint. The 19 comments which accompanied the first-round ratings were distributed similarly: The Brophy and Good review received five positive comments and one negative comment; the Doyle review received no positive comments, three comments which contained both positive and negative elements, and three negative comments; and the Rosenshine and Stevens review received one positive, five mixed, and one negative comment. In the second round, there were only five comments, made by three panelists; in general, the second-round comments seem to reinforce those made in the first round.

It had been recognized that a possible outcome of Phase I might be

Table 2

Ratings from Phase I Delphi Panel

	minimally reflective					highly reflective
BROPHY AND GOOD	1	2	3	4	5	6	7
first round mean = 5.8	L				B C M Y	A M X Z	
second round mean = 5.1				B	A C L M Y Z	N X	
DOYLE	1	2	3	4	5	6	7
first round mean = 3.2		N X Z	C L	B M Y	A		
second round mean = 2.9		X Y	A B C M N Z	L			
ROSENSHINE AND STEVENS	1	2	3	4	5	6	7
first round mean = 3.8		L	M N X	A C Z		B Y	
second round mean = 3.6			L M N X Z	A B C		Y	

Note: Panelists' ratings are represented by an identifying letter, rather than a tally, to allow readers to compare first and second round ratings of individual panelists. Letters were assigned to panelists as follows:

A elementary reading
B elementary social studies
C elementary special education

L secondary English
M secondary math
N secondary special education

X supervision, elementary
Y supervision, secondary
Z supervision, district level

the selection of more than one review for use in Phase II, and Phase II procedures had been designed to accommodate that possibility. Because the Brophy and Good review was the only one with a preponderance of positive comments and the only one rated at or above the midpoint by all of the panelists, and because these ratings contrasted considerably with the ratings for the other two reviews, Brophy and Good's "Teacher Behavior and Student Achievement" was the only review selected for use in Phase II.

Phase II

Panelists. Individuals participating in Phase II of this study considered items from the Test of Professional Knowledge and classified them according to the findings included in "Teacher Behavior and Student Achievement" (Brophy and Good, 1986). Each of the 12 Phase II panelists had at least three years' experience as a classroom teacher, had served as a cooperating teacher to undergraduate students in full-time student teaching placements, and had received in-service training in teacher effectiveness. Nine of the panelists had participated in a 6-week, 15-hour Master Teacher Development Course, which was designed to strengthen the skills of cooperating teachers and included recommendations drawn from research on teaching; the remaining three panelists had a minimum of nine hours of inservice education in research-based teacher effectiveness strategies.

Procedure. Two panels were assembled, each with six members; panelists teaching in the same school were assigned to different panels. Each panel considered one-half, or 52, of the 104 items on the Test of Professional Knowledge included in a released form of the NTE. Each Phase II panelist received a packet containing:

-a detailed set of directions

-a copy of "Teacher Behavior and Student Achievement," the review of teacher effectiveness research by Brophy and Good (1986) selected in Phase I of this study

-an additional copy of the section of the review titled "Summary and Integration of the Findings," reproduced word for word and in the same order as the original, but organized to make the panelists' task easier by grouping on a single page all of the research findings included in a single category. Also, each paragraph in the summary was numbered.

-the 52 test items

-a set of eight demonstration items, with possible classifications and comments, and

-a stamped, return envelope.

Panelists were asked to examine the review thoroughly, then to read each test item with its keyed answer and to classify it in terms of the research findings included in the review. The available classifications were: strongly supported by research, moderately supported by research, unrelated to the research cited in the review, moderately contradictory to research, or strongly contradictory to research. If the item was classified as either supported by or contradictory to research, the panelist was asked to identify the relevant research by writing the number appearing next to the related paragraph in the research summary. Panelists were also invited to comment on their classifications.

Panelists sent their item classifications and comments to the Delphi coordinator, who tabulated the responses and returned them to the panelists. In the second round of Phase II, each panelist received:

-a brief summary of the first-round data and directions for classifying the items a second time, in light of the feedback from other panelists

-a second copy of the test items, with the classifications, related paragraphs, and comments panelists supplied for each item during the first round, and with space for second-round classifications and comments, and

-a reaction sheet, which asked panelists to identify from among the test items classified as "supported by research", the items for which knowledge of the research findings was important in selecting the keyed option.

After completing the second round classifications and the reaction sheets, panelists returned them by mail to the Delphi coordinator.

Results. For the purposes of this study, an item was considered "supported by research" if at least four of the six panel members classified the item as either strongly or moderately supported by a particular research finding. Using this criterion, 20 of the 104 items (19%) can be considered supported by research. Nine of these 20 items (9% of the total items) were judged by a majority of panelists to be items for which knowledge of the related research finding was important in selecting the keyed option. The research findings cited by the panelists and the test items with which they are associated are presented in Table 3. The research finding cited most often (about success rates and academic learning time) and the five items associated with it are presented in Table 4.

Seventy-one items (68%) can be considered "unrelated to the research cited:" 56 items were classified by all six panel members as unrelated and 15 were classified unrelated by four or five panelists.

None of the items were classified "moderately contradictory" or "strongly contradictory." Although some panelists used the moderately contradictory classification, none of the items were judged contradictory to research by four or more panelists.

Panelists also identified, from among the items they classified as

Table 3
Research Findings Associated with TPK Items Classified "Supported by Research"

Research Findings	TPK Items: Supported by research (a)	TPK Items: Research important in selecting keyed option (b)
Quantity and Pacing of Instruction		
Opportunity to Learn/Content Covered		
Role Definition/Expectations/Time Allocation		
Classroom Management/Student Engaged Time	1	1
Consistent Success/Academic Learning Time	5	2
Active Teaching		
Whole Class vs. Small Group vs. Individualized Instruction	3	3
Giving Information		
Structuring	3	2
Redundancy/Sequencing		
Clarity		
Enthusiasm		
Pacing/Wait-Time		
Questioning the Students		
Difficulty Level of Questions	1	
Cognitive Level of Questions	1	
Clarity of Question		
Post-Question Wait-Time		
Selecting the Respondent		
Waiting for the Student to Respond		
Reacting to Student Responses		
Reactions to Correct Responses	3	1
Reacting to Partly Correct Responses		
Reacting to Incorrect Responses		
Reacting to "No Response"		
Reacting to Student Questions and Comments	1	
Handling Seatwork and Homework Assignments	1	
Context-Specific Findings		
Grade Level		
Student SES/Ability/Affect	1	
Teacher's Intentions/Objectives		
Other		

(a) Total number of items classified "supported by research" and linked with research finding. (b) Of those classified "supported by research", total number for which knowledge of the research finding was judged important in selecting the keyed option.

Table 4
Consistent Success/Academic Learning Time

Research Finding(a)

Consistent Success/Academic Learning Time. To learn efficiently, students must be engaged in activities that are appropriate in difficulty level and otherwise suited to their current achievement levels and needs. It is important not only to maximize content coverage by pacing the students briskly through the curriculum, but also to see that they make continuous progress all along the way, moving through small steps with high (or at least moderate) rates of success and minimal confusion or frustration. If lessons are to run smoothly without loss of momentum and students are to work on assignments with high levels of success, teachers must be effective in diagnosing learning needs and prescribing appropriate activities. Their questions must usually (about 75% of the time) yield correct answers and seldom yield no response at all, and their seatwork activities must be completed with 90-100% success by most students. (Such high success rates should not be taken as suggestive of instructional overkill or assignment of pointless busywork. Appropriate seatwork will extend knowledge and provide needed practice. It will also be do-able, however, because it is pitched at the right level and because students have been prepared for it. Thus the high success rates result from effort and thought, not mere automatic application of already overlearned algorithms). Continuous progress at high rates of success, carried to the point that performance objectives can be met smoothly and rapidly, is especially important in the early grades and whenever students are learning basic knowledge or skills that will be applied later in higher-level activities. (Brophy and Good, 1986, p. 360-361)

Associated TPK Items

* 50A. Of the following, the most important element in the effective use of individualized instruction is:

(A) effective communication between a student's parents and teachers
(B) the establishment of appropriate evaluation standards
+ (C) accurate diagnosis and prescription of learning
(D) the availability of attractive instructional materials
(E) the identification of possible information resources

Classifications (b): Strongly supported - 2; Moderately supported - 4

Note: Items marked with asterisks are those for which knowledge of the related research was judged important in selecting the keyed option.

(a) The research about consistent success/academic learning time is discussed in three paragraphs in the "Summary and Integration of the Findings" section of the Brophy and Good review. However, in classifying TPK items, panelists cited only the first paragraph of the discussion, which is reproduced here. (b) Indicates number of panelists assigning each classification.

Table 4, continued
Consistent Success/Academic Learning Time

6B. Which of the following should receive consideration by a teacher who is preparing a reading list from which students select required reading materials?

I. Student interests
II. Reading level of the selections
III. Availabilty of the selections
IV. Community resources.

(A) I and II only
(B) I and IV only
(C) I, II, and III only
(D) II, III, and IV only
+ (E) I, II, III, and IV

Classifications: Strongly supported - 1; Moderately supported - 4; Unrelated - 1

* 5A. Research indicates that in classrooms where effective teaching and learning occur, the teacher is likely to be doing which of the following consistently?

(A) Gearing instruction to the typical student at a given grade level
(B) Carefully grouping students at the beginning of he school year and making sure that these groups remain the same throughout the year
(C) Identifying the affective behaviors that students are likely to exhibit at a given level of development
(D) Working diligently with students to make sure that each learns all the material planned for the class for the year
+ (E) Pacing instruction so that students can move ahead when they are able to or receive extra help when they need it

Classifications: Strongly supported - 3; Moderately supported - 1;
Related to a different research finding - 2

13A. Good instructional planning is built around the idea that what learners will learn is most often determined by

(A) what they should know
(B) what their teacher knows
+ (C) how and why they learn
(D) who does the teaching
(E) what parents and administrators desire

Classifications: Moderately supported - 5; Related to a different research finding - 1

38A. A policy of equal educational opportunity obligates the teacher in which of the following ways?

(A) Every child must be taught the same things.
(B) All children must be treated alike.
(C) Instruction must exclude use of multi-cultural learning materials.
(D) Every class must have a proportionate minority population.
+ (E) Instructional strategies must be adapted to the individual.

Classifications: Moderately supported - 4; Unrelated - 1;
Related to a different research finding - 1

supported by research, the items for which knowledge of the related research findings was important in selecting the keyed option. There were no items identified by all six panelists as meeting this criterion. One item was identified by five panelists, and eight items were identified by four panelists as items for which knowledge of the related research was important in selecting the correct answer.

The results of Phase II would have differed slightly if a different criterion had been used, specifically, if an item were considered "supported by research" when three, rather than four, panelists classified the item as either strongly or moderately supported by a particular research finding. Using the "three of six" criterion, 28 rather than 20 items (27% rather than 19%) would be considered supported by research. Of the eight additional items, five were paired with the research finding about consistent success/academic learning time -- the finding cited most often under the previous criterion, and two items were paired with other research findings already cited.

Limitations of the Study

Conclusions that may be drawn from this study are subject to certain limitations imposed by the Delphi method, and by the use of a single form of the NTE and a single review of research in Phase II.

All Delphi studies are subject to an a priori limitation: the judgment achieved through the Delphi method represents a consensus among experts, but there is no guarantee that it represents the "best" judgment. In addition, in this study, Phase II panelists classifying items as supported by research were limited to the research findings included in the Brophy and Good review. If the Phase I panel had chosen a different review or an additional review, there would have been differences in the item

classifications. However, at least one researcher's informal analysis of the content of the TPK has yielded results that correspond to those generated in this study: Darling-Hammond (1986) concluded that "less than 10% of over 100 questions required knowledge of theory, research, or facts pertaining to teaching and learning" (p. 20).

Discussion

In commenting on the extent to which knowledge of the research findings is important in selecting the keyed answer for NTE items, one panelist said, "...the questions seem to be of the 'common sense' variety..."; another panelist stated, "I can see how someone with good general knowledge (higher SAT scores) and good test-taking ability would be able to do well without exposure to educational research." In these comments, the panelists echo critics who have suggested that the NTE Test of Professional Knowledge measures something other than teachers' professional knowledge. Evidence provided by Andrews, Blackmon, and Mackey (1980), Miller, Poggio, and Glasnapp (1987), Loadman (1987), Lovelace and Martin (1984), Pitcher (cited in Wilson, 1986), and Weber and McBee (1987) supports Nelsen's (1985) conclusion that

> performance variations may be largely attributable to factors such as general intelligence, scholastic aptitude, overall academic achievement, and multiple-choice test item reasoning skills, rather than to the extent of instruction or mastery of particular domains of the curriculum, such as professional education. (p. 1066)

To the extent to which the test measures such factors as general intelligence rather than teachers' professional knowledge, the TPK can be regarded as lacking in educational importance. The educational importance of a test can be questioned when the test measures something unimportant or

fails to measure something important (Cronbach, 1971). Insofar as it measures scholastic aptitude and overall academic achievement, the TPK might be regarded as educationally unimportant, not because these factors are unimportant in initial teacher certification decisions, but because other measures of them (e.g., SAT or GRE scores) already exist as part of the educational record of nearly every applicant for teacher certification. In addition, judging from the results of this study, the TPK fails to measure important aspects of teachers' professional knowledge. Although the TPK contains 20 items judged by panelists to be related to 10 research findings, the panelists estimated that for 11 of those items (55%), knowledge of the related research was not important in selecting the correct answer. In covering part of the professional knowledge base for teachers with items requiring only good general knowledge and/or common sense, the TPK misses opportunities to measure some important aspects of teachers' professional knowledge. For example, consider the following item:

> 3A. Each term, a teacher provides book lists from which the students choose books about which they will write book reports. Some topics seem to appeal more to girls and others appeal more to boys. The teacher could best help the students find books that will most likely appeal to them by doing which of the following?
>
> (A) Listing all of the books by reading level
> * (B) Listing all of the books by subject
> (C) Grouping all of the books by length
> (D) Making up one list for boys and another list for girls
> (E) Making up one list of books by male authors and another by female authors (ETS, 1984, p. 109)

This item was classified by four of the six panelists as moderately supported by the research on structuring. One panelist commented that "Paragraph 14 - structuring - does deal with the skills of presenting information and structuring techniques. There is a slight connection -- this listing could be an advance organizer." The discussion of structuring in the summary of the Brophy and Good review mentions, in addition to advance organizers, "overviews, review of objectives; outlining the content

and signaling transitions between lesson parts; calling attention to main ideas; summarizing subparts of the lesson as it proceeds; reviewing main ideas at the end;...using organizing concepts, analogies, [and]...rule-example-rule patterns" (p. 362). While there may be debate about whether the item relates to advance organizers, it seems clear that answering it correctly does not depend on teachers' professional knowledge about structuring.

Although there are two other TPK items considered supported by research and linked to the research on structuring, they too measure knowledge of structuring only rudimentarily. Those two items are particularly noteworthy because they were the only items in this study classified by all six panel members as strongly supported by research. The items are presented below.

12A. A fourth grade class is going to visit a museum for the first time. In order to prepare the students to learn from the experience, the teacher should do which of the following?

I. Give the pupils a set of questions about the exhibits in an effort to focus their attention during the visit.
II. Tell pupils about museums--what they are and why people visit them.
III. Have a lesson about some of the exhibits pupils will see on the trip.
IV. Tell the pupils the field trip will be a test of their ability to practice good citizenship.

(A) I only
(B) II only
(C) I and IV only
* (D) I, II, and III only
(E) II, III, and IV only (ETS, 1984, p. 111)

40A. Which of the following, if given to high school students at the beginning of a new course, is an example of an advance organizer?

(A) A list of books required to do the supplementary reading
* (B) An overview of the course that includes objectives and assessment criteria
(C) An essay assignment to determine levels of writing skill in the

class
(D) A lecture about discipline and behavior standards in the classroom
(E) A reading test to determine the students' ability to read material in the content field (ETS, 1984, p. 116)

Both items were judged by four of the six panelists as items for which knowledge of the research finding was important in selecting the keyed option. The first item, 12 A, measures an aspect of structuring that is not closely related to the specific and complex elements of structuring described by Brophy and Good as part of lesson presentation. The second item, 40 A, was cited by Darling-Hammond (1986) to demonstrate the very elementary nature of even those few TPK items that do require "knowledge of theory, research, or facts pertaining to teaching and learning" (p. 20).

The research finding on structuring and the three TPK items associated with it deserve careful consideration because only one research finding was associated with more items (Consistent Success/Academic Learning Time, associated with five items) and because no other items were judged to be as strongly supported by a research finding. However, it would appear that even these items cannot be cited as evidence of the educational importance of the TPK, because they measure less important aspects of the topic and fail to measure more important elements. As Darling-Hammond (1986) has said, the TPK is "limited...by the scarcity of important teaching questions answerable in multiple-choice formats; the questions with clear, correct answers are not very profound" (p. 21). The items associated with structuring may demonstrate what Bracey (1987) has described as the tendency for minimum competency tests to emphasize trivial objectives at the expense of more difficult aspects of the curriculum which may be harder to assess.

A lack of evidence supporting the educational importance of the TPK raises questions of construct validation. Although NTE publications provide evidence of content validation and do not discuss construct validation, some critics (e.g., Madaus and Pullin, 1987; Nelsen, 1985) have argued that

construct validation is essential for a test such as the NTE. Standards for Educational and Psychological Testing (AERA, APA, and NCME, 1985) includes a description of test validation as a process that requires evidence of content-, criterion-, and construct-related validity. It is interesting to note that shortly before the release of the revised Core Battery, an NTE staffer co-authored an article which included the statement that "convincing arguments place construct validity at the heart of questions involving test interpretation and use, thus making it an imperative adjunct to any future research or operational effort to improve the NTE" (Rosner and Howey, 1982, p. 7).

It seems reasonable to assume that a measure of a construct involving teachers' professional knowledge would necessarily include items designed to measure the ability to apply knowledge derived from research on teacher effectiveness. Because a majority of the items classified as related to research could be answered correctly without knowledge of the research, the construct underlying performance on the TPK would seem to involve this component of teachers' professional knowledge only minimally and to be related instead to factors that are not specific to teachers' professional knowledge, such as "general intelligence, scholastic aptitude, overall academic achievement, and multiple-choice test item reasoning skills" (Nelsen, 1985, p. 1066).

The extent to which the TPK measures such factors while neglecting some elements of teachers' professional knowledge can be viewed as troubling in light of the role of the TPK in shaping teacher education curricula. In 22 states, applicants for teacher certification must pass the NTE; in some states, teacher education programs with a specified percentage of graduates who do not pass the NTE are threatened with loss of their state approval (Goodison, 1986). Faculty members in teacher education programs want their

graduates to earn certification, so it is inevitable that the NTE will have an influence on teacher education curricula. The extent of this influence is suggested in a New York State Education Department memorandum reporting the results of a survey of efforts made by colleges to aid members of minority groups in passing the NTE. The memorandum lists nine activities reported by colleges, including "Revision of the curriculum to reflect knowledge necessary to pass the NTE, especially in courses devoted to the teaching-learning process" and "Offering a two-credit course in preparation for the NTE" (Van Ryn, 1987, p. 4).

Acknowledging the power of the NTE to shape teacher preparation curricula, Shulman (1987) has argued that initial teacher certification tests

> must become <u>tests worth teaching for.</u> The traditional criteria of reliability and validity are no longer sufficient. As long as assessments drive instruction, assessment designers have a moral obligation to create instruments that correspond to appropriate images of excellent professional preparation and practice. (p. 44)

Conclusions

It is unfortunate that the NTE Test of Professional Knowledge, adopted by 22 states in the wake of the educational reform movement, may have the unintended effect of impeding other parts of that movement. Considerable effort has been devoted to building an understanding of the professional nature of teachers' work, and to countering a public perception that reasonably competent adults do not need special preparation to become effective teachers. In offering a test of professional knowledge on which only 9% of the items were judged to require knowledge of research in teacher

effectiveness, the NTE may be seen as reinforcing the notion that teachers' professional knowledge is little more than good general knowledge and common sense. This simplistic view of teaching may lead to a superficial definition of the professional knowledge base and threaten efforts to enhance the professional status of teaching.

Recently, ETS acknowledged the limitations of the current NTE and announced plans to replace it with a "radically different" test that will be available to states by 1992 (Olson, 1988, p. 1). The new test is expected to differ from the current NTE in its use of advances in technology to allow for adaptive testing, and in the timetable for test administration. Unlike the current NTE, which can be completed in a single day at any point before certification, the new (and still unnamed) test will be administered at three separate stages in a teacher's career: after the sophomore year, a computerized diagnostic battery will assess basic skills; at the end of the teacher-education sequence, a paper-and-pencil test will measure knowledge of content and pedagogy; and after a substantial practice teaching experience or internship, a performance test will assess the ability to teach a given content area in a classroom setting. The performance test may be supplemented with computer simulation exercises and with portfolios that document a teacher's work (Dwyer, 1988).

Given that the president of ETS, Gregory Anrig, described the test as "radically different" and called the test development process a "full court press" representing a "high risk" for ETS (Olson, 1988, p. 27), the new test could be a significant departure from earlier revisions. However, it seems likely that the portion of the new test designed to assess teachers' knowledge of pedagogy will continue to resemble the current Test of Professional Knowledge. Like the TPK, the new test will be administered when a prospective teacher has completed the teacher education sequence but

before he or she has substantial classroom experience. Until recently, ETS described it as a paper-and-pencil test using a multiple choice format like the current professional knowledge test (Dwyer, 1988); however, it now appears that ETS is exploring the inclusion of some constructed response items, such as short-answer items, along with the multiple-choice items (Fiero, 1990).

Some critics contend that no objective test is likely to yield an adequate measure of teachers' professional knowledge: "In general, the state of the art does not permit objective tests for directly measuring higher order thinking skills, problem solving strategies, and metacognitive abilities involved in tasks such as teaching" (Fredericksen and Collins, 1989, p. 29). Advances in technology, such as the interactive videodisc, may soon be applied to assessment methods and allow for improved measurement of complex knowledge and skills. But for several reasons, including cost considerations and lack of equal access to computers in many areas where the test will be used, teachers' professional knowledge may continue to be measured primarily using traditional assesement methods. And, to extend a caution advanced by Renfrow and Cromrey (1990), such changes in format could be cosmetic, enhancing face validity only.

If the new test is to be modeled to some extent on the current TPK, the data generated in this study may provide the test writers with some new perspectives on the content of a test of teachers' professional knowledge. Darling-Hammond (1986) has observed that many of the TPK items require examiness to "choose a teaching technique in response to short scenarios that give insufficient information to make a truly reasoned judgement. ...a thoughtful, honest, and knowledgeable teacher would in most cases have to answer, 'It depends.'" (p.46). In describing challenges facing the new NTE, Dwyer (1989) has identified "the need to contextualize the assessment and to

bring it closer to specific teaching situations" (p.36). However, as long as the new test of pedagogy remains a paper-and-pencil test using primarily multiple-choice items, test developers may wish to consider including more items that test knowledge of research findings at the knowledge level rather than the application level.

More than half of the items on the current TPK are at the application level (ETS, 1984) and include brief descriptions of classroom situations. ETS's decision to test teachers at the application level is understandable, particularly in light of test users' demands that licensure exams demonstrate job relevance. However, there may be insurmountable difficulties inherent in using multiple-choice items to measure teachers' ability to apply professional knowledge, particularly knowledge derived from research on teaching. Research seldom, if ever, yields direct rules for practice. Effective teaching is highly context-sensitive, and recommendations drawn from research that are effective in one context may be ineffective or counterproductive in another setting. (Darling-Hammond's comment that the correct answer to most TPK items is "It depends" reflects this reality.) Although findings from research on teaching cannot provide prospective teachers with a recipe to follow in any given classroom situation, they can help teachers analyze classroom events and formulate plans for action that are based on more than intuition. A primary value of the findings from teacher effectiveness research is that, as Gage (1985) has said, "(they) give teachers something to reason with and about."

Describing teaching contexts in sufficient detail to allow accurate measurement of the ability to apply knowledge of research on teaching to classroom situations seems to be a task that is unlikely to be accomplished using traditional assessment methods. (However, it may be possible to do this with interactive videodiscs; the technology currently available seems

well suited to representing classrooms in all their vitality and complexity, directing examinees' attention to one aspect of the classroom situation and posing a question about it, then presenting subsequent items based on examinees' responses to earlier ones.) Since multiple-choice items are likely to continue to be a mainstay of the successor to the TPK, it may be most appropriate to use them to determine if examinees can answer factual questions about key concepts such as academic learning time, structuring, or wait time. In using multiple-choice items to test knowledge of research findings at the knowledge level only, item writers can avoid application items that suggest there is a single correct response to a given classroom situation, as well as items that are so general that they can be answered correctly by examinees who do not have knowledge of the underlying concept.

Clearly, testing knowledge of research on teaching at the knowledge level only is not a satisfactory long term solution. In preparing good application items, developers of the new NTE may want to consider variations on traditional items and new strategies for item validation. Norris (1989), in discussing the development of tests of critical thinking, has suggested strategies that may be adapted successfully to testing teachers' professional knowledge.

In a modification of an objective test, Norris has asked examinees to justify, orally or in writing, their answers to multiple-choice items. When generating application items, NTE developers might consider similar two-part items. A standard multiple-choice question might be followed by a second question, also using a multiple-choice format, that asks for justification for the first response, as a means of determining the data or reasoning an examinee used in selecting one option over another. Presumably, an examinee who selected the right option for the wrong reason

would be penalized.

Norris (1989) also has identified strategies that may be useful in validating application items, specifically, asking examinees to think aloud while working on items and asking them to describe how they used particular pieces of information presented in the item in selecting an answer. Data gathered in this way can be used in modifying application items when, for example, it appears that examinees have selected an option other than the key, despite having recognized and used the relevant research, or when examinees do not give evidence of using research findings but still are able to arrive at the keyed response to an item designated as requiring application of knowledge derived from research on teaching.

Recommendations for Further Research

Additional research related to this topic might include: variations on the present study; a survey of the ways research on teaching currently is presented in teacher preparation curricula, which could lead to new TPK items reflective of exemplary practice in teacher education; and the exploration of ways to assess the professional knowledge of beginning teachers using formats other than multiple-choice test items.

Variations on this study might generate results that are different from those reported here and suggest other interpretations of how findings from research on teacher effectiveness are represented on the Test of Professional Knowledge. For example, the use of a different or additional review of research in Phase II could lead to results suggesting more or less congruence between findings from research on teacher effectiveness and TPK items, or to different understandings of the extent to which certain findings are represented. Another valuable variation would be the presentation of TPK items without the key. In the current study, all items

were presented with the keyed option marked by an asterisk; it might be useful to see if not directing panelists to the intent of an item in that way would lead to differences in the extent of agreement among panelists.

Stallings (1984) has noted that findings from research on teaching were disseminated first to practicing teachers through inservice education, and only later to preservice teachers in teacher education programs. This lag in dissemination may help account for the relatively weak representation of findings from research on teaching on the NTE Test of Professional Knowledge, because validation studies (conducted primarily in the early 1980s) involved comparison of the test items with the content of teacher preparation programs. Given the increased attention to the importance of including findings from research on teaching in teacher education curricula (for example, the recommendations for reform of teacher education issued by the National Commission for Excellence in Teacher Education [1985]), it is reasonable to expect that the content of teacher education programs has been modified considerably since the earliest validation studies were conducted. A survey of the ways teacher educators present findings from research on teaching and assess mastery of that portion of the body of professional knowledge could lead to the identification of exemplars of outstanding practice for the benefit of both teacher educators and TPK item writers.

It is likely that a meaningful test of a beginning teacher's professional knowledge will measure the ability to apply learning theory, knowledge of child or adolescent development, and recommendations drawn from research on teaching in ways that are context-sensitive, that is, in ways that respond to differences in student ability level and educational setting and are appropriate to particular subject areas and grade levels. This will be extremely difficult to accomplish within the limits of a paper-and-pencil test using primarily a multiple-choice format. Lee Shulman (1987) and his

colleagues on the Teacher Assessment Project (TAP) are currently developing prototypes for assessing the competence of experienced teachers who will seek voluntary professional certification through the National Board of Professional Teaching Standards. ETS should be commended for seeking to apply some of the ideas from the TAP to assessment for entry-level teachers, and encouraged to continue to explore alternate formats for the measurement of teachers' professional knowledge in ways that are meaningful and representative of classroom practice.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1985). Standards for Educational and Psychological Testing. Washington, D.C.: American Psychological Association.

Andrews, J. W., Blackmon, C. R., & Mackey, J. A. (1980). Preservice performance and the National Teacher Exams. Phi Delta Kappan, 61, 358-359.

Berliner, D. C. (1984). The half-full glass: A review of research on teaching. In P. L. Hosford (Ed.), Using what we know about teaching (pp. 51-77). Alexandria, VA: Association for Supervision and Curriculum Development.

Bracey, G. W. (1987). Measurement-driven instruction: Catchy phrase, dangerous practice. Phi Delta Kappan, 68, 683-686.

Brophy, J., & Good, T.L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 328-375). New York: Macmillan.

Clark, C. M. (1984). Research on teaching and the content of teacher education programs: An optimistic view. (Occasional Paper No. 75). East Lansing, MI: Michigan State University, Institute for Research on Teaching.

Cronbach, L. J. (1971). Test validation. In R. L. Thorndike, (Ed.), Educational Measurement (2nd ed., pp. 443-507). Washington, DC: American Council on Education.

Darling-Hammond, L. (1986). Teaching knowledge: How do we test it? American Educator, 46, 18-21.

Doyle, W. (1982). Research on classroom contexts: Toward a knowledge base for policy and practice in teacher education.In G. A. Griffin & H. Hukill (Eds.), Alternate Perspectives for Program Development and Research in Teacher Education (pp. 75-94).Austin, TX: University of Texas at Austin, Research and Development Center for Teacher Education. (ERIC Document Reproduction Service No. ED 223-578)

Doyle, W. (1986). Classroom organization and management. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 392-431). New York: Macmillan.

Dwyer, C.A. (1988, November). A new generation of tests for teacher licensure. Paper presented at the annual meeting of the Northeastern Research Association, Ellenville, NY.

Dwyer, C.A. (1989). A new generation of tests for licensing beginning teachers. New Directions for Teacher Assessment: Proceedings of the 1988 ETS Invitational Conference (pp. 29-38). Princeton, NJ: Educational Testing Service.

Educational Testing Service. (1984). A guide to the NTE Core Battery Tests. Princeton, NJ: Author.

Egbert, R.L. (1984). The role of research in teacher education. In R.L. Egbert and M.M. Kluender (Eds.), Using research to improve teacher education: The Nebraska consortium. (Teacher Education Monograph No. 1, pp. 9-21). Washington, DC: ERIC Clearinghouse on Teacher Education.

Fiero, D. (1990). Working description of Stage II. Working papers toward a new generation of teacher assessments. Princeton, NJ: Educational Testing Service.

Frederiksen, J. R. & Collins, A. (1989). A systems approach to educational testing. Educational Researcher, 18, 27-32.

Gage, N. L. (1978). The scientific basis of the art of teaching. New York: Teachers College Press.

Gage, N. L. (1985). Hard gains in the soft sciences: The case of pedagogy. Bloomington, IN: Phi Delta Kappa, Center on Evaluation, Development, and Research.

Good, T. L. (1983). Research on classroom teaching. In L. S. Schulman & G. A. Sykes (Eds.), Handbook of Teaching and Policy (pp. 42-80). New York: Longman.

Goodison, M. (1986). Teacher testing: Improving the outcomes. Paper presented at the annual meeting of the American Educational Research Association, San Francisco.

Hersh, R. M. (1982). What makes some schools and teachers more effective. In D. C. Corrigan (Ed.), The future of teacher education: Needed research and practice (pp. 1-9). College Station, TX: Texas A & M University, College of Education. (ERIC Document Reproduction Service No. ED 225 997)

Hindman, S. E. & Polsgrove, L. (1988). Differential effects of feedback on preservice teacher behavior. Teacher Education and Special Education, 11, 25-29.

Hosford, P. L. (1984). The art of applying the science of education. In P. L. Hosford (Ed.), Using what we know about teaching (pp. 141-161). Alexandria, VA: Association for Supervision and Curriculum Development.

Hunter, M. (1984). Knowing, teaching, and supervising. In P. L. Hosford (Ed.), Using what we know about teaching (pp. 169-192). Alexandria, VA: Association for Supervision and Curriculum Development.

Kluender, M.M. (1984). The Nebraska consortium: Using research to improve teacher education. In R.L. Egbert and M.M. Kluender (Eds.), Using research to improve teacher education: The Nebraska consortium. (Teacher Education Monograph No. 1, pp. 1-8). Washington, DC: ERIC Clearinghouse on Teacher Education.

Linstone, H. A., & Turoff, M. (Eds.). (1975). The Delphi method: Techniques and applications. Reading, MA: Addison-Wesley.

Loadman, W.E. (1987, April). Use of the NTE as a measure of competence for teacher certification. Paper presented at the annual meeting of the American

Educational Research Association, Washington, DC.

Lovelace, T. & Martin, C. E. (1984). The revised National Teacher Examinations as a predictor of teachers' performance in public school classrooms. Lafayette, LA: University of Southwestern Louisiana. (ERIC Document Reproduction Service No. ED 251 416)

Madaus, G. F. & Pullin, D. (1987). Teacher certification tests: Do they really measure what we need to know? Phi Delta Kappan, 69, 31-38.

Miller, M. D., Poggio, J. P., & Glasnapp, D. R. (1987). Teachers' professional knowledge: Are we measuring acquired skills or common knowledge? Journal Of Personnel Evaluation in Education, 1, 57-68.

National Commission for Excellence in Teacher Education. (1985). A call for change in teacher education. Washington, DC: American Association of Colleges for Teacher Education.

Nelsen, E. A. (1985). Review of NTE programs. In J. V. Mitchell (ed.), The ninth mental measurements yearbook (pp. 1063-1066). Lincoln, NE: Buros Institute of Mental Measurements of the University of Nebraska - Lincoln.

Norris, S. P. (1989). Can we test validly for critical thinking? Educational Researcher, 18, 21-26.

Olson, L. (1988, November 2). 'Different' tests of teaching skill planned by firm. Education Week, pp.1, 27.

Pipho, C. (1986). States move reform closer to reality. Phi Delta Kappan, 68, K1-K8.

Rosenshine, B. & Stevens, R. (1986). Teaching functions. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 376-391). New York: Macmillan.

Rosner, F. C. & Howey, K. R. (1982). Contruct validity in assessing teacher knowledge: New NTE interpretations. Journal of Teacher Education, 33, 6, 7-12.

Rudner, L. M. (1987). Questions and answers concerning teacher testing. In L. M. Rudner (Ed.), What's happening in teacher testing: An analysis of state teacher testing practices (pp. 3-8). Washington, D.C.: US Department of Education, Office of Educational Research and Improvement.

Shulman, L. S. (1987). Assessment for teaching: An initiative for the profession. Phi Delta Kappan, 69, 38-44.

Smith, B. O. (1985). Research bases for teacher education. Phi Delta Kappan, 66, 685-690.

Smith, D. C. (1982, May). The content in teacher education programs. Paper presented at the Conference on the Future of Teacher Education: Needed Research and Practice, College Station, TX. (ERIC Document Reproduction Service No. ED 217-029)

Smith, D. C. (1983, February). PROTEACH: An extended preservice teacher

preparation program. Paper presented at the annual meeting of the American Association of Colleges for Teacher Education, Detroit. (ERIC Document Reproduction Service No. ED 232 976)

Stallings, J. A. (1984). Implications from the research on teaching for teacher preparation. In R.L. Egbert and M.M. Kluender (Eds.), Using research to improve teacher education: The Nebraska consortium. (Teacher Education Monograph No. 1, pp. 128-146). Washington, DC: ERIC Clearinghouse on Teacher Education.

Trentham, L., Lauderdale, W., Miller, E., & Rudder, C. (1986, April). Relationships among the Alabama Initial Teacher Certification Test scores, college variables, and classroom performance ratings. Paper presented at the annual meeting of the National Council on Measurement in Education, San Francisco.

U.S. Department of Education. (1986). What works: Research about teaching and learning. Washington, D.C.: Author.

U.S. Department of Education. (1987). What works: Research about teaching and learning. (2nd ed.). Washington, D.C.: Author.

Van Ryn, M. (1987, December 8). [Results of a memorandum concerning attraction and recruitment of minorities to teaching.] Memorandum from the Assistant Commisioner for Higher Education Services, State Education Department, University of the State of New York.

Walberg, H. J. (1986). Syntheses of research on teaching. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed., pp. 214-229). New York: Macmillan.

Weber, L. J., & McBee, J. K. (1987, April). A challenge to the validity of the Professional Knowledge Test of the Core Battery of the National Teachers Examination [sic]: A teacher licensing problem. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC.

Wilson, A. J. (1986, April). Historical issues of validity and validation: The National Teacher Examinations. Paper presented at the annual meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 270 503)

Wyne, M. D. & Stuck, G. B. (1982). Time and learning: Implications for the classroom teacher. Elementary School Journal, 83, 68-75.